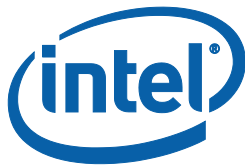


Delivering Low Cost High IOPS VM Datastores Using Nexenta* and the Intel[®] SSD Data Center S3500 Series

Solutions Blueprint

November 2013

Revision 1.0



INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Configurations: 1. NexentaStor* appliance: Intel® Dual-Processor 2x Intel® Xeon® E5-2690, Intel® S2600CP Xeon® Motherboard, 128 GB of RAM in 16x 8GB, 24x 2.5-inch drive 2U Server, 2x LSI 9207-8i HBA, 12x 2.5-inch 800GB Intel® SSD DC S3500 Series SATA SSD , 1x 2.5-inch 400GB Intel® SSD DC S3700 Series SATA SSD, 1x Intel x520-DA2 Network Adapter. 2. ESXi Hosts: Intel Dual-Processor 2x Intel® Xeon® E5-2690, Supermicro® Super X9DRFR Xeon® Motherboard, 128 GB of RAM in 16x 8GB, 1x 2.5-inch Boot/Swap, 1x Intel x520-DA2 Network Adapter

Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.

Intel does not control or audit the design or implementation of third party benchmark data or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmark data are reported and confirm whether the referenced benchmark data are accurate and reflect performance of systems available for purchase.

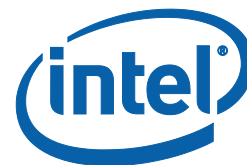
For more information go to <http://www.intel.com/performance>.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: <http://www.intel.com/design/literature.htm>

Intel and Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

*Other names and brands may be claimed as the property of others.

Copyright© 2013 Intel Corporation. All rights reserved.



Contents

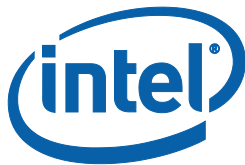
1	Introduction.....	5
1.1	Audience and Purpose	5
1.2	Prerequisites	5
1.3	Solution Summary	5
1.4	Terminology	6
1.5	Reference Documents.....	7
2	Solution Overview.....	8
2.1	Existing Challenges	8
2.2	Intel Solution	8
2.3	Solution Impact	8
3	Solutions Implementation	10
3.1	Nexenta* Product Overview	11
3.2	NexentaStor* Additional Features	12
4	Test Configuration and Plan	13
4.1	Hardware Setup and Configuration.....	13
4.2	Software Setup and Configuration.....	14
4.3	Software Architecture and Network Configuration Diagram	15
5	Walk Through Setup and Configuration.....	16
5.1	Basic Installation – Nexenta*	16
5.2	Network Configuration.....	16
5.3	Optimize ZFS for 4KiB Block Devices	17
5.4	Configure Storage and Features.....	21
5.5	Configure NFS	23
5.6	VMware* NFS Configuration	26
5.7	Calculating SSD Endurance in RAID.....	28
5.8	Configuration and Setup Results	29
5.8.1	Results.....	29
5.8.2	Conclusion	29

Figures

Figure 4-1. System Architecture Diagram	14
Figure 4-2. Network Configuration Diagram	15

Tables

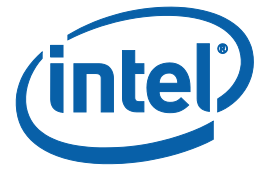
Table 5-1. Sample Endurance Calculation--5-Disk RAID5 Set DC S3500 Series.....	29
--	----



Revision History

Document Number	Revision Number	Description	Revision Date
329591-001	1.0	Initial release.	November 2013

§



1 Introduction

1.1 Audience and Purpose

The purpose of this document is to give IT professionals instruction on how to construct a purpose-built high-IOPS/low-cost Storage Area Network (SAN) using off-the-shelf Intel components, the Nexenta* software stack, and VMware* ESXi 5.1. This software-based storage solution is specifically targeted at highly dense virtual environments that interleave VM I/O resulting in a mostly random I/O demand, also known as the I/O Blender. This document steps through basic installation and setup of hardware (HW), Nexenta* software (SW), and connection of NFS datastores at the VMware host system to NFS shares on a Nexenta* server.

In addition, this document explains a best known method for hardware and NFS setup in Nexenta* software for optimal performance as a VM datastore using Intel® 10Gb Ethernet controllers and Intel® Solid-State Drive Data Center S3500 Series (Intel® SSD DC S3500 Series) products.

Additional details on Nexenta* may be found in the Nexenta* installation manual and administration guide located at:

<http://info.nexenta.com/rs/nexenta/images/NexentaStor-InstallationGuide-3.1.x.pdf>

Additional details concerning VMware* setup may be found in the vSphere* 5.1 administrator's guide and related documents located at:

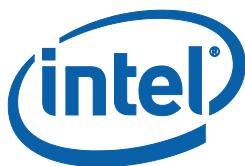
<http://www.vmware.com/support/product-support/vsphere-hypervisor.html>. In addition to installation guides and manuals, many helpful tips may be found in both Nexenta's* and VMware's* community blog forums and support sites.

1.2 Prerequisites

This document assumes the reader has prior knowledge of hardware and systems, basic Linux* configuration and installation, and advanced knowledge of VMware* ESXi and vSphere* 5.1 interface—specifically storage configuration.

1.3 Solution Summary

The solution described as configured within this document is capable of supplying virtual machines (VMs) with 100K random 4K block mixed read/write I/O operations per second (IOPS) for an approximate cost of \$30K while using only 2 units of rack space and up to 650 watts of power. Industry background and the adoption of 10Gb Ethernet (10GbE) networking have allowed flexibility and efficiency in the data center. Businesses employing 10GbE can replace many 1Gb twisted pair cables connecting an individual server with only two 10GbE cables, saving costs on both structured cabling and switch ports without sacrificing performance. Using NFS* shares, iSCSI*, or Fibre Channel over Ethernet (FCoE*) protocols with 10GbE allows for removal of parallel storage networks and reduces cabling and switching complexity. This document describes how to construct an NFS storage solution using off-the-shelf Intel® Xeon® processor based servers, Intel® 10GbE Network Adapters, Intel® SSD DC S3500



Series, and the Nexenta* software stack. Building an equivalent IOPS capable solution from a simple array of 15K RPM high performance 2.5" disks would require approximately 500 disks, 40U of rack space, 7.5 Kilowatts of power, and cost roughly \$125K for the disks alone. Similarly, a SAN solution capable of producing this type of sustained random I/O could cost over \$1 per IOPS from a typical storage vendor. In both cases as long as the target storage footprint does not exceed the proposed 8TB solution presented here, it offers superior performance to the competing solution in IOPS per dollar and overall cost and upkeep.

Note: 15K RPM drives cost and performance numbers based on an industry average of 200 random IOPS per 15K RPM drive at 15W of power.

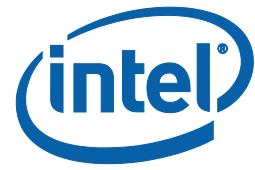
General assumptions for calculations				
Drive Type	IOPS	Cost	Power Consumption	Size
15K RPM SAS	200	\$300	15W	600GB
Intel® SSD DC S3500 Series	75000	\$900 ¹	5W	800GB

Note:

1. Cost numbers based on suggested retail pricing of Intel® SSD DC S3500 800GB drive as of September 2013.

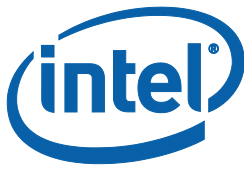
1.4 Terminology

Term	Description
Copy-on-write (CoW)	A computer programming optimization strategy.
HA	VMWare's* High Availability
Hypervisor	Software, also known as a virtualization manager or VMM (virtual machine monitor) that manages virtual machines
KIB	Kibibyte, 1024 bytes (vs Kilobyte which is 1,000 bytes)
IOPS	Input/output operations per second
JBOD	Non-RAID storage architecture where each drive is accessed directly as a standalone device. Acronym stands for "Just a Bunch Of Drives."
LUN	Logical Unit Number
NFS	Network File System
PID	Physical Identification
SAN	Storage Area Network
TCO	Total Cost of Ownership
VID	Vendor Identification
VM	Virtual Machine



1.5 Reference Documents

Document	Document Number/Location
Nexenta* Quick Start Installation Guide	http://info.nexenta.com/rs/nexenta/images/doc_3.1_nexentastor-quickstart-3-1.pdf
Nexenta* Installation Guide	http://info.nexenta.com/rs/nexenta/images/NexentaStor-InstallationGuide-3.1.x.pdf
vSphere* 5.1 Administration Guide	http://www.vmware.com/support/product-support/vsphere-hypervisor.html
Intel® Ethernet Converged Network Adapter Home Page	http://www.intel.com/content/www/us/en/network-adapters/converged-network-adapters.html
Intel® SSD DC S3500 Series Home Page	http://www.intel.com/content/www/us/en/solid-state-drives/solid-state-drives-dc-s3500-series.html
Intel® Xeon® E5 Family Home Page	http://www.intel.com/content/www/us/en/processors/xeon/xeon-processor-5000-sequence.html



2 Solution Overview

Intel® Solid-State Drives balance performance, endurance, and lower solution cost to meet virtualization and cloud demands.

2.1 Existing Challenges

Pressure to move toward cloud. Cloud computing offers a multitude of benefits. Balancing the enterprise private cloud with public offerings, IT organizations must increase internal density to compete with the public cloud's economy of scale.

Many VMs per datastore. As internal density increases, more and more VMs reside in single large storage area network (SAN) -based logical drives (LUNs). This interleaves I/O and produces, from the SAN's perspective, a random workload.

Doing more for less. All IT organizations are driven by budget constraints. Any solution that requires a smaller data center footprint, uses less power, and leverages hardware resources more efficiently helps accommodate these constraints.

2.2 Intel Solution

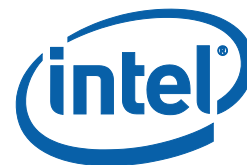
Intel® SSD DC S3500 Series. Deploying a software based NAS solution to handle many VMs using the Intel® SSD DC S3500 Series can provide a low-power, low-cost, high IOPS solution relative to existing alternatives.

2.3 Solution Impact

100k Random I/O Operations per Second (IOPS). The Intel® SSD DC S3500 Series Nexenta* based software SAN delivers 100K random 4kB IOPS at a 90/10 read/write ratio, the equivalent of roughly 500 15K RPM high-performance hard disk drives (HDDs), at a much smaller power budget of 650 watts. This configuration, as detailed in sections 4.1 and 4.2 is capable of driving hundreds of concurrent VMs.

Lower costs. When compared with the cost of powering the hundreds of 15K RPM high-performance HDDs needed to produce 100K IOPS, this Intel® SSD-based solution consumes 650 watts versus roughly 7.5 kilowatts -- 11x less power. When compared with the average SAN at \$1 per random IOPS for 100K sustained workloads; this \$30K solution comes in at only \$0.30 per random IOPS, a 70% cost savings. In addition, the 2U footprint is smaller than most comparable performing SANs, allows for an additional 12 disks of growth, requires less cooling budget, and reduces overall complexity for simplified management.

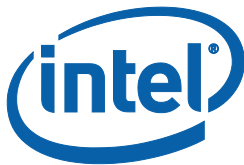
Note: Cost includes small percentage for server chassis, board, controllers, and network.



General assumptions for calculations					
Drive Type	IOPS Target	Drives Required	Cost	Power Consumption	\$/IOPS
15K RPM SAS ¹	100K	500	\$150,000	7.5kW	\$1.50
Intel [®] SSD DC S3500 Series ²	100K	12	\$30,000 ¹	650W	\$.30
<i>Summary</i>	100K	2U vs. 40U	<20% Capex	11x less power	<80% cost/IOPS

Notes:

1. 15K RPM high-performance drives cost and performance numbers based on an industry average of 200 random IOPS per 15K RPM drive at 15W of power.
2. Cost numbers based on suggested retail pricing of Intel[®] SSD DC S3500 Series 800GB drive as of September 2013.



3 Solutions Implementation

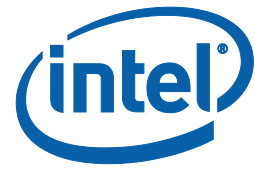
Building a private cloud is a large and complicated task. One of the main challenges revolves around shared storage and SAN. A decade ago, servers were built with 3-8 disks in a Redundant Array of Independent Disks or RAID set which serviced the operating system and application. In addition, high performance storage needs were addressed with SAN and the only items placed in high-cost SAN were databases. Fast forward to 2013, the typical server has only one or two local disks with a virtualization layer or hypervisor and all the virtual machines sitting in SAN. In essence, we have inverted the disk ratio where 10 years ago we had 3-8 disks for an OS/app stack now we have 3-8 OS/app stacks per disk in a SAN logical unit number (LUN).

One byproduct of this consolidation to VMs is a phenomenal ramp in the randomness of disk I/O. For example, say you have a *single* virtual machine (VM) connected to a SAN with 14x disks in a RAID array as your VM datastore. This configuration behaves exactly like a dedicated server and the disk I/O profile of the OS/app stack is identical. Now, you introduce a *second* virtual machine and have combined the two OS/app disk I/O profiles as they read and write to the same VM datastore and your disk I/O becomes 50% random when viewed from the perspective of the SAN. Continue to add VMs to this same datastore and *with 10 or more active VMs*, the I/O pattern is almost completely random as the individual VMs' I/O streams interleave. Random I/O does a few unsightly things to the typical SAN: It renders caching algorithms useless for predicting I/O, and with caching ineffective you revert to the least common denominator of about 200 IOPS per physical disk running at 15K RPM. Should IOPS demand from any particular VM reach 2.8K IOPS, we have exhausted the capability of our theoretical 14x disk array and every other VM in this datastore degrades because of a single noisy neighbor.

Note: The typical VM datastore at 2-4TB in size can easily host 50-100 VMs at 40GB storage footprint per VM which makes I/O contention commonplace.

The solution, examined in this document, is to construct a SAN which is composed entirely of Solid State Drives (SSDs). Still more expensive per drive than traditional spinning disks (HDDs) in most cases, the SSD has no moving parts and no seek time which makes random I/O a perfect fit for these drives. SSDs cost more in \$ per Gigabyte (GB), however, since they generate phenomenal volumes of IOPS per drive the cost per IOPS is dramatically less than traditional disks. For instance, a 800GB Intel® SSD DC S3500 Series drive can produce 75K random read IOPS at an approximate list price of \$940, which yields 79 IOPS per dollar. Whereas a 15K RPM 600GB HDD can produce 200 random read IOPS at \$300, or 0.67 IOPS per dollar. The bottom line in virtualized and cloud infrastructure is that consolidating VMs produces an uptick in random IOPS, random IOPS devalue caching algorithms, and hard disk drives cannot deliver these random IOPS efficiently. SSDs on the other hand continue to decline in cost per GB, excel at providing random I/O, run on approximately one third the power, and have now evolved with enterprise-class high endurance features and consistency features.

It seems a logical conclusion that for high-density VM environments such as the ones found in enterprise IT data centers, an SSD-based SAN is an optimal solution. The solution examined in this blueprint shows an SSD based solution using Intel®



components and Intel's enterprise class Intel® SSD DC S3500 Series Multi-Level Cell (MLC) NAND which allows for up to 5 years of continuous writes to the drive at 1x drive overwrites every 3 days.

3.1 Nexenta* Product Overview

NexentaStor* is a fully featured NAS/SAN software platform with enterprise class capabilities that address the challenges presented by ever-growing datasets. NexentaStor* continues to build on its position as a superior, open source, multi-platform alternative to vertically integrated and proprietary storage offerings.

Certified on a wide range of industry standard x86-based hardware, it provides users the flexibility to select their storage operating system independently from the hardware it runs on. This enables users to select the best components for their need.

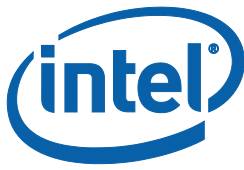
NexentaStor* helps organizations enjoy significant benefits including higher workload throughput and improved management capabilities. Using NexentaStor* as a storage foundation enables organizations to implement high-performance; cost-effective data storage solutions with features that include inline compression, unlimited snapshots, cloning and support for highly available storage. This strategy reduces IT complexity and delivers compelling storage economics.

NexentaStor* 4.0 is a full featured storage operating system built on community supported IllumOS* (currently Open Solaris* in 3.1.4), enabling organizations to deploy an incredibly fast file system without being locked into a single proprietary storage hardware supplier.

Based on the ZFS File System and through partnerships with the open-source community, NexentaStor* benefits from an exchange of product development and support activities that collectively allow for greater resources than any single storage vendor can provide. At its core is open source code and built around its Nexenta* proprietary IP code, seamlessly integrated and packaged for commercial deployment and use.

NexentaStor* can create shared pools of storage from any combination of storage hardware, including solid state drives. It also delivers end-to-end data integrity, integrated search and inline virus scanning.

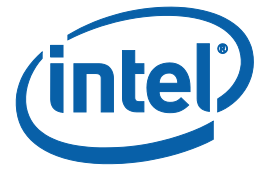
The power of NexentaStor* is extended by the use of its licensed features. These include namespace cluster, high availability clusters, OST for Symantec NetBackup* and Nexenta* Target FC.* These also provide integration enhancements for management and administration with other 3rd Party products such as virtualization suites. Incidentally, these optional features can be installed without restarting NexentaStor* or upgrading the entire system.



3.2 **NexentaStor* Additional Features**

Additional features include:

- **Highly Scalable File System:** ZFS-based technology provides a highly scalable and flexible 128-bit file system
- **End-to-End Integrity:** Because of the self-healing nature of the ZFS file system, users need not worry about silent data corruption
- **Unlimited Storage Capacity:** Designed for large-scale growth with unlimited snapshots and copy-on-write clones
- **Largest File Size Commercially Available:** While most legacy solutions limit the size of the file system, there are no such limitations with ZFS
- **Unified Appliance:** Support for NAS (NFS, CIFS WebDAV, FTP) and SAN (iSCSI and FC) means flexibility
- **Data De-duplication and Native Compression:** Inline data de-dupe and compression reduces the use and expense of primary storage
- **Hybrid Storage Pools:** Exploit multiple layers of cache to accelerate read and write performance
- **Heterogeneous Block and File Replication:** Asynchronous replication for easy disaster recovery
- **Block-level Mirroring:** Setup remote backups and disaster recovery at offsite locations
- **Simplified Disk Management:** Manage disks and JBODs using either command line or the web-based interface provided in the NexentaStor* Management Viewer (NMV)
- **Total Product Safety:** Proactive, automated alert mechanisms send alerts when configurations deviate from the IT defined user-based policy



4 Test Configuration and Plan

4.1 Hardware Setup and Configuration

It is important that all hardware utilized is listed on the Nexenta* HSL, Hardware Supported List. <http://www.nexenta.com/corp/support/support-overview/hardware-supported-list>

Hardware setup used for NexentaStor*:

- Intel® Dual-Processor 2 Intel® Xeon® E5-2690
- Intel® S2600CP Xeon® Motherboard
- 128 GB of RAM in 16x 8GB
- 24 2.5-inch drive 2U Server
- 2 LSI 9207-8i HBA
- 12 2.5-inch 800GB (SATA) Intel® SSD DC S3500 Series
- 1 2.5-inch 400GB (SATA) Intel® SSD DC S3700 Series
- 1 Intel® x520-DA2 Network Adapter

Hardware configuration used for NexentaStor*:

Configuration consists of 12 Intel® SSD DC S3500 Series, connected to 2 HBAs; 6 drives each. This is commonly referred to as a channel per drive configuration as there is no expander being used. Each drive has a direct SAS connection to the HBA and balance the I/O to/from the PCIe* BUS for maximum performance. Four more drives could be added, two to each HBA, to increase capacity and performance if necessary, but is not covered in this guide. Additionally, there are 12 more drive bays in the solution used, so expanding with an additional SAS controller and 8 more drives attached to that controller is also a possibility.

The Solid-State Drives are configured in two software RAID Z1 volumes, one per HBA. While RAID Z1 is comparable to a traditional hardware RAID 5, RAID Z1 uses copy-on-write rather than the slower read/modify/write or write hole to update blocks.

Figure 4-1. System Architecture Diagram

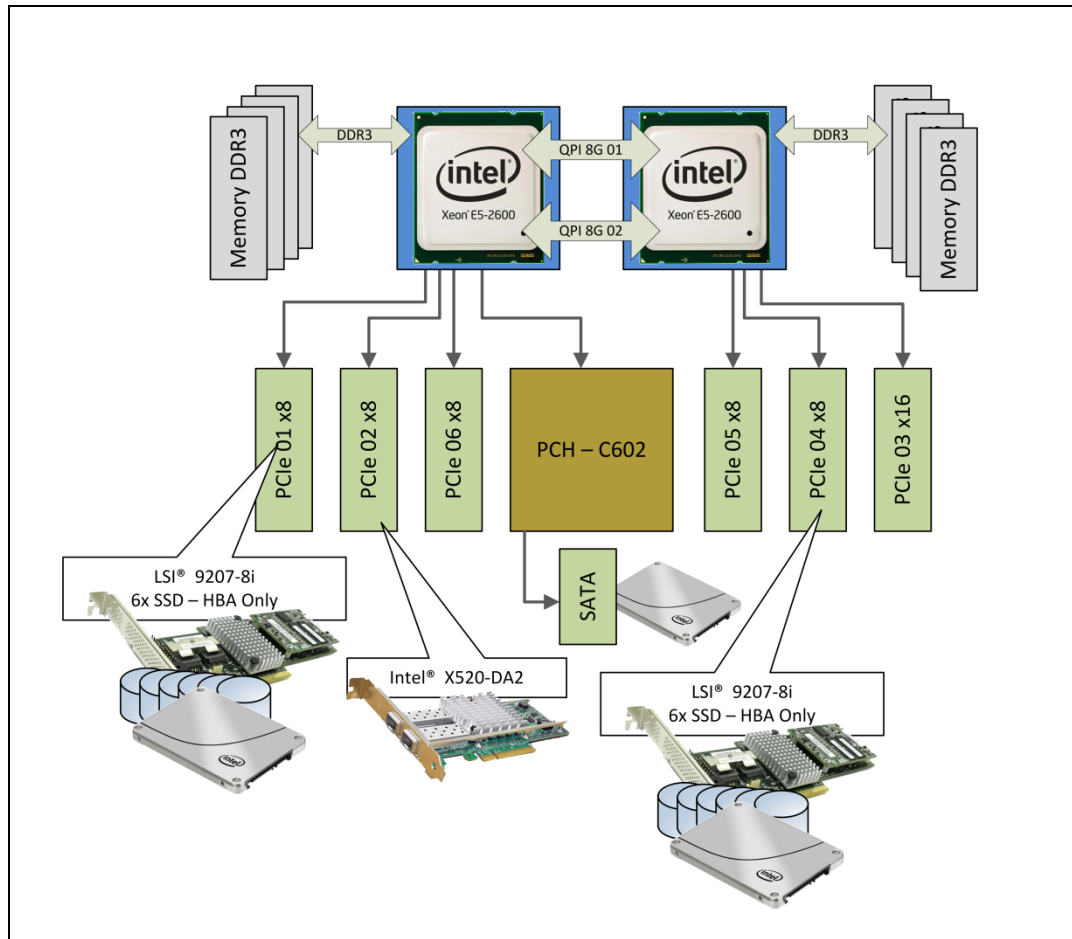


Figure 4-1. System Architecture Diagram shows the PCIe to controller relationships between the two 8-port SAS HBAs (controllers), the Intel® x520-DA2 network controller, and the local SATA controller for the VMware® ESXi hypervisor. This configuration was chosen for the ability of the system with two controllers to handle the I/O produced by 12 SSDs. The configuration spreads I/O across both processors in this dual-socket system. Further performance gains are possible using more/many SAS controllers, and additional 10GbE network cards to spread the I/O load even further across multiple CPUs and I/O paths, however these expanded configurations were not tested in the scenario presented in this paper.

4.2 Software Setup and Configuration

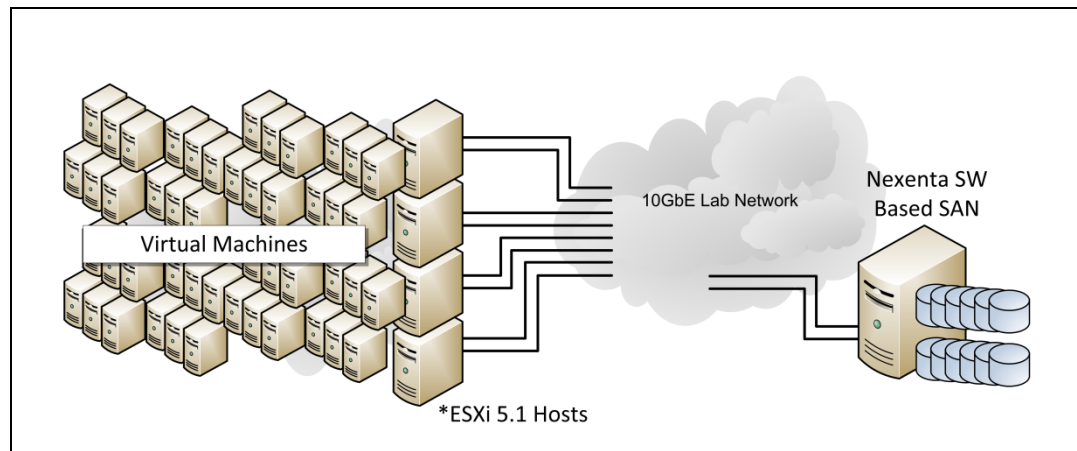
Software used in the high density VM datastore SAN and host configuration and testing.

- SAN/NAS – NexentaStor* version 3.1.4
- Virtualization/Cloud – VMware* ESXi and vCenter server version 5.1 Update 2
- VMs – Windows* Server 2008 R2 w/current updates and McAfee VSE* 8.8 patch 2
- IO load generation – Iometer* v2006.07.27 available from iometer.org

4.3 Software Architecture and Network Configuration Diagram

The diagram below shows the relationship between the 48 VMs loaded with Windows* Server 2008 R2 and Iometer, the four hosts loaded with ESXi 5.1 U2 used to house those 48 VMs, and the 10GbE lab network across which storage traffic to the NFS based VM datastores occurs on the NexentaStor* server.

Figure 4-2. Network Configuration Diagram



5 Walk Through Setup and Configuration

5.1 Basic Installation – Nexenta*

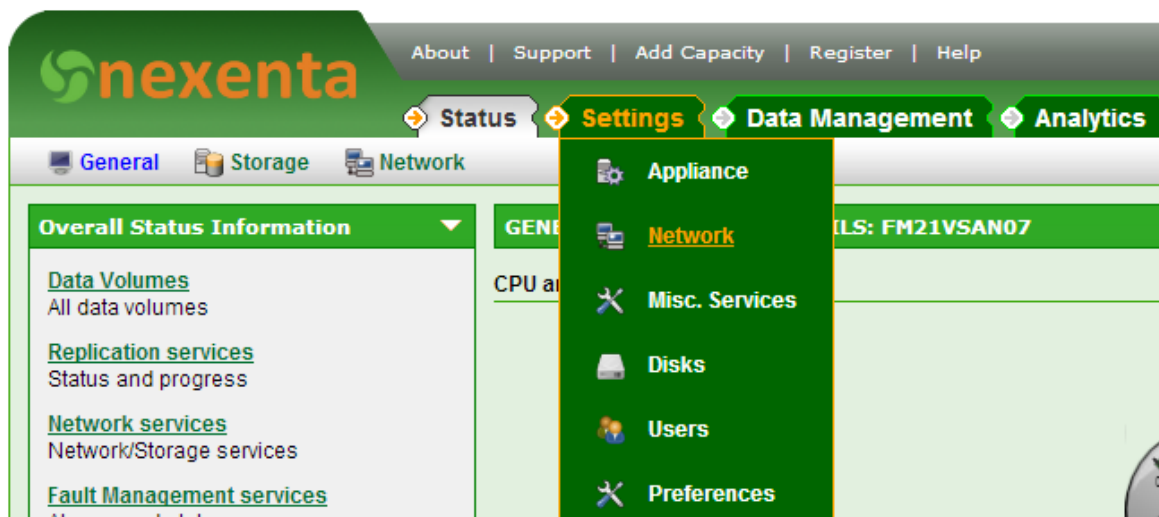
The basic installation and configuration of the NexentaStor* 3.1.4 software stack is not covered in this document. Refer to the NexentaStor* QuickStart Guide and Installation Guide to install and perform the basic setup of the software.

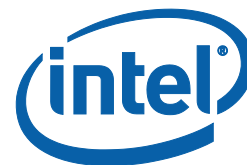
http://info.nexenta.com/rs/nexenta/images/doc_3.1_nexentastor-quickstart-3-1.pdf

5.2 Network Configuration

In order for the storage requests to be load balanced to the NexentaStor* appliance, it is important to pay special attention to the network configuration. Requests must traverse both network ports to alleviate potential network bottlenecks. Placing each 10GbE port on a different subnet will enable TCP/IP routing decisions to be made. If these ports were placed on the same subnet, there would be no way to enable each 10GbE port to be used efficiently. It is true that LACP port channels could be used on the network, but that would take the load balancing decisions away from the systems and place it at the switching layer.

1. Open a web browser and navigate to the NexentaStor* appliance `http://<IP or Hostname>:2000/status/general/`
2. Navigate to **Settings** tab and select **Network**





Walk Through Setup and Configuration

3. Verify that you have configured the two 10 GbE network adapters on different subnets.

SUMMARY NETWORK SETTINGS				
Network Interfaces:				
Interface	Type	Configuration	Primary	Actions
igb0	physical	Configured as 10.80.143.17/255.255.255.0 with mtu 1500	<input checked="" type="radio"/>	<input type="text"/>
igb1	physical	Unconfigured	<input type="radio"/>	<input type="text"/>
igb2	physical	Unconfigured	<input type="radio"/>	<input type="text"/>
ixgbe0	physical	Configured as 172.16.5.17/255.255.255.0 with mtu 1500	<input type="radio"/>	<input type="text"/>
ixgbe1	physical	Configured as 172.16.4.17/255.255.255.0 with mtu 1500	<input checked="" type="radio"/>	<input type="text"/>

Note: 1GbE and 10 GbE adapters are listed as *igb* and *ixgbe* respectively.

5.3 Optimize ZFS for 4KiB Block Devices

Until recently, the majority of disk drive vendors used a block or sector size of 512 Bytes and subsequently, most filing systems are optimized for 512B blocks. Many SSD vendors today are optimizing around sector sizes of 4096 Bytes or 4KiB. These drives still support 512B blocks but do so differently. 512B sectors are now emulated or listed as the logical block size, and 4KiB sectors are listed as the physical sector size. This is because each 4KiB sector will actually contain eight smaller 512B sectors on the disk. The operating system can still access these 512B sectors but will do so with a performance penalty for write transactions. Running disk benchmark utilities such as Iometer will reveal these performance barriers with most 4KiB drives at the sub 4KiB transaction size.

ZFS can be tuned to operate at a 4KiB sector size and increase performance. To do so, the NexentaStor* must be told which drives are 4KiB based and the drives in this paper are indeed based on the larger sector size. Modifying the system file `sd.conf`, and adding lines to identify the Intel drives as 4KiB will help to accomplish this.

4. Using PuTTY* or secure shell tool; SSH* to the NexentaStor* appliance. Log in as admin then SU to root.

```
admin@fm21vsan07: /export/home/admin
login as: admin
Using keyboard-interactive authentication.
Password:
Last login: Wed Jul 3 09:48:47 2013 from jphubbar-mob11.

* * *

CAUTION: This appliance is not a general purpose operating system:
managing the appliance via Unix shell is NOT recommended. Please use
management console (NMC). NMC is the command-line interface (CLI) of
the appliance, specifically designed for all command-line interactions.
Using Unix shell without authorization of your support provider may not
be supported and MAY VOID your license agreement. To display the
agreement, please use the following NMC command:

show appliance license agreement

admin@fm21vsan07: ~$ su
Password:
root@fm21vsan07: /export/home/admin#
```

NOTE: Logging in as Root will bring you to the Nexenta* Management Console (NMC). The changes being made will need to be done through a command shell, not the NMC.



- ```
echo "::walk sd_state | ::grep '!=0' | ::print struct sd_lun un_sd |
::print struct scsi_device sd_inq | ::print struct scsi_inquiry inq_vid
inq_pid" | mdb -k
```

[illegible]

The output of the command referenced in step 5 should read as follows:

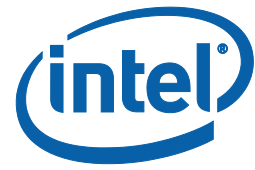
```
inq_vid = ["ATA "]
inq_pid = ["INTEL SSDSA2BZ20"]
inq_vid = ["ATA "]
inq_pid = ["INTEL SSDSC2BB80"]
```

"ATA INTEL SSDSA2BZ20"

```
physical-block-size:4096
```

throttle-max:32

**disksort:false**



## Walk Through Setup and Configuration

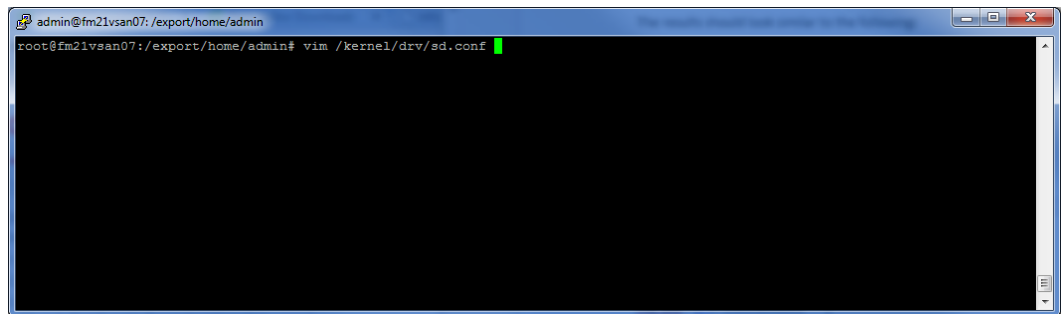
By default the OS optimizes for HDD seek distance. Intel® SSD Series do not have any moving parts, thus optimizing for seek distances hurts SSD performance.

**cache-nonvolatile:true**

The Intel® SSD DC S3700 Series and Intel® SSD DC S3500 Series drives are protected with PLI circuitry (Power Loss Imminent), that commit all unwritten data in the transfer buffer to storage in the event of a power loss. Cache flushing with PLI protected SSDs can lead to performance hits and is unnecessary.

6. Edit /kernel/drv/sd.conf with VIM

```
vim /kernel/drv/sd.conf
```



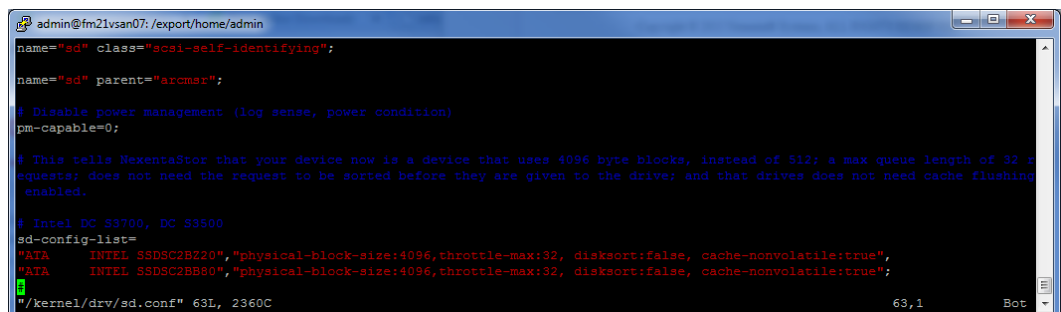
- a. Navigate to the end of the file and add the following lines to the configuration. Adding these comments can make the file easier to read for future modifications.

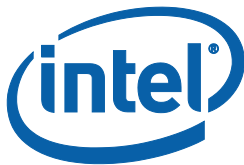
```
Intel DC S3700 SSD, Intel DC S3500 SSD
```

```
sd-config-list=
```

```
"ATA INTEL SSDSC2BZ20", "physical-block-size:4096,throttle-max:32, disksort:false, cache-nonvolatile:true",
```

```
"ATA INTEL SSDSC2BB80", "physical-block-size:4096,throttle-max:32, disksort:false, cache-nonvolatile:true";
```





This configuration tells NexentaStor\* that both the Intel® SSD DC S3700 Series and Intel® SSD DC S3500 Series have a 4KiB physical block size, max queue depth of 32, disable sorting of seek distance requests, and disable cache flushing.

Also pay attention the comma after the first entry for the Intel® SSD DC S3700 Series, and the semicolon after the last entry.

When presenting multiple disk types, a comma separated list ending with a semicolon is necessary.

Sample sd.conf sd-config-list entry:

```
sd-config-list=

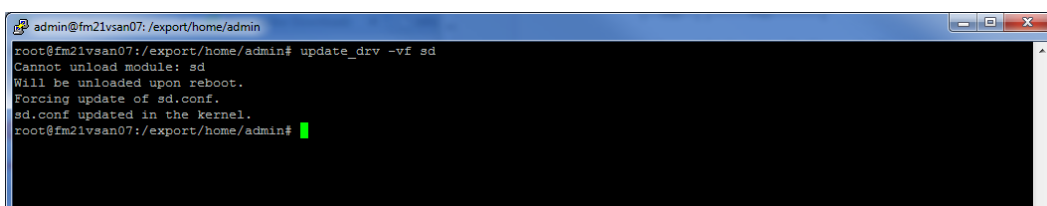
 "Disk Drive", "Options",

 "Disk Drive", "Options",

 "Disk Drive", "Options";
```

7. When finished modify the file, update the OS with the following command.

```
update_drv -vf sd
```



```
admin@fm21vsan07:/export/home/admin
root@fm21vsan07:/export/home/admin# update_drv -vf sd
Cannot unload module: sd
Will be unloaded upon reboot.
Forcing update of sd.conf.
sd.conf updated in the kernel.
root@fm21vsan07:/export/home/admin#
```

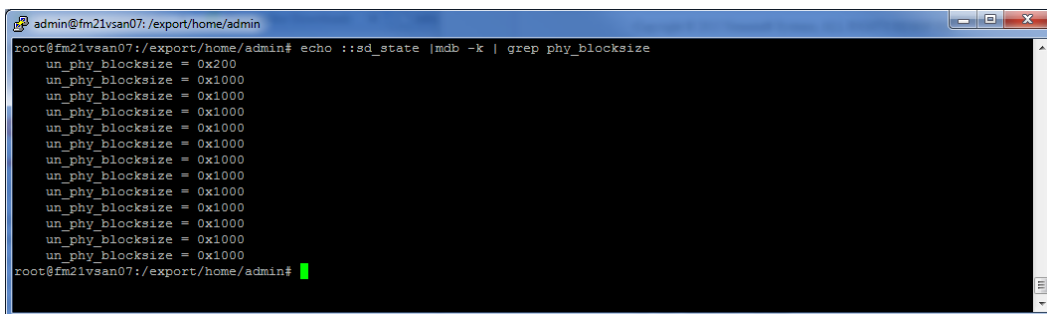
Pay attention only to the last two lines. The OS has been correctly updated if these following lines are present.

```
Forcing update of sd.conf

Updated in the kernel.
```

8. Verify the SSDs have been updated to use a 4KiB block size.

```
echo ::sd_state | mdb -k | grep phy_blocksize
```



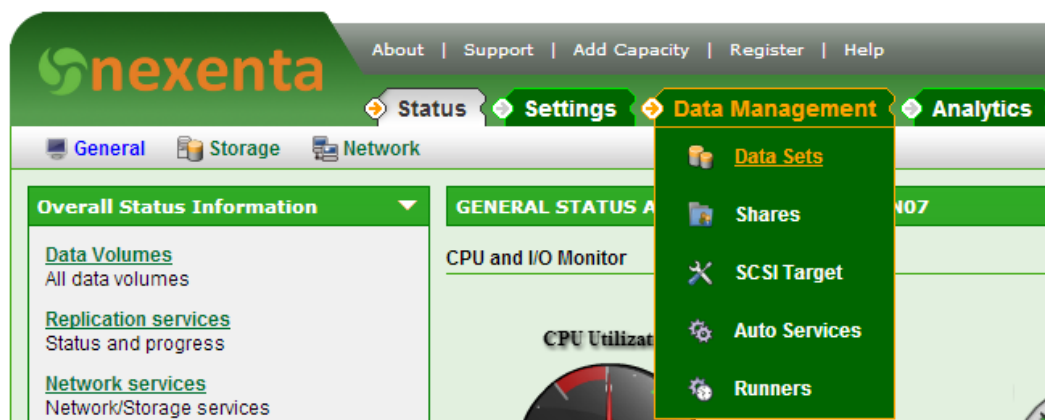
```
admin@fm21vsan07:/export/home/admin
root@fm21vsan07:/export/home/admin# echo ::sd_state | mdb -k | grep phy_blocksize
un_phy_blocksize = 0x200
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
un_phy_blocksize = 0x1000
root@fm21vsan07:/export/home/admin#
```

**Note:** Block sizes of 0x200 and 0x1000 represent 512B and 4KiB respectively.

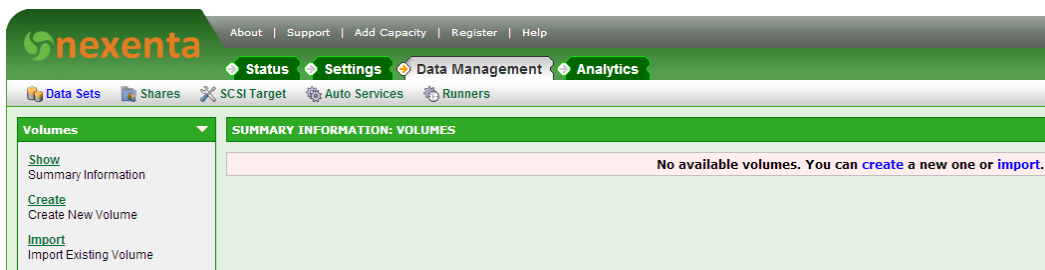
## 5.4 Configure Storage and Features

Now that the storage devices have been optimized it's time to concentrate pooling them together.

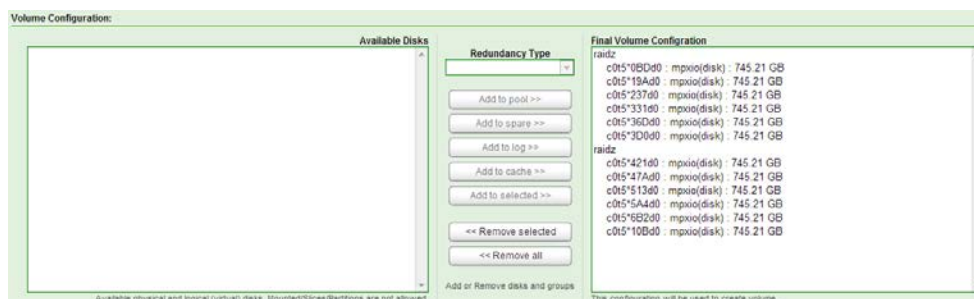
1. Open a web browser and navigate to the NexentaStor\* appliance.
2. Navigate to the **Data Management** tab, and select **Data Sets**.

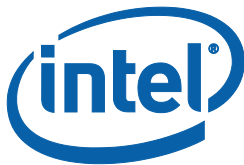


3. Navigate to the **Volumes** field and select **Create**.



4. Create a ZPOOL.
  - a. Select 6 disks under Available Disks.
  - b. Select **RAID-Z1** (single parity) from the Redundancy Type.
  - c. Click the **Add to Pool >>** Button.
  - d. Select the remaining 6 disks under **Available Disks**.
  - e. Select **RAID-Z1** (single parity) from the Redundancy Type.
  - f. Click the **Add to Pool >>** button.





## Walk Through Setup and Configuration

5. Enter a name and description
  - a. Enable deduplication and select **OK** to the warning.
  - b. Leave the other values at their defaults and select the **Create Volume** button.

**Volume Properties:**

|                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Name</b>                                  | <input type="text" value="zpool"/><br><small>Volume name must begin with a letter and can only contain alphanumeric characters (a-z, A-Z, 0-9) in addition to the following three special characters: underscore (_), hyphen (-) and period (.)<br/>Volume name has the following restrictions: the beginning sequence c[0-9] is not allowed; the name log is reserved; a name that begins with mirror, raidz, or spare is not allowed because these name are reserved.<br/>In addition, volume name must not contain a percent sign (%).</small> |
| <b>Description</b>                           | <input type="text" value="Intel SSD zPool"/><br><small>Optional volume description. Maximum length is 255 characters.</small>                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>Deduplication</b>                         | <input type="text" value="on"/><br><small>Controls the deduplication option for the volume. If enabled, it will optimize use of duplicate copies of data. Default is off.</small>                                                                                                                                                                                                                                                                                                                                                                 |
| <b>Compression</b>                           | <input type="text" value="on"/><br><small>Controls the compression algorithm used for this dataset. Default is "on". Setting compression to "on" uses the lzjb compression algorithm. The lzjb compression algorithm is optimized for performance while providing decent data compression. Currently, "gzip" is equivalent to "gzip-6".</small>                                                                                                                                                                                                   |
| <b>Autoexpand</b>                            | <input type="text" value="off"/><br><small>Controls automatic pool expansion when the underlying LUN is grown.</small>                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>Sync</b>                                  | <input type="text" value="standard"/><br><small>Controls synchronous requests (standard - ensure all synchronous requests are written to stable storage; always - every file system transaction will be written and flushed to stable storage by system call return; disabled - synchronous requests are disabled). Default is standard.</small>                                                                                                                                                                                                  |
| <b>Force creation</b>                        | <input type="checkbox"/><br><small>Forces use of (virtual or physical) disks (LUNs) even if they appear to be in use.</small>                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <input type="button" value="Create Volume"/> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |

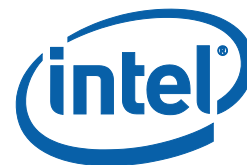
- c. Provide Administrative credentials and select login.

**NMV Login**

Permission denied. You do not have sufficient permissions to complete add/create request. Please click the browser's "back" button, or log in as a different user.

**User Name:**

**Password:**



## Walk Through Setup and Configuration

The zpool has been created and we are returned to data sets view under data management.

| SUMMARY INFORMATION: VOLUMES  |                              |         |           |         |          |             |        |        |        |        |  |
|-------------------------------|------------------------------|---------|-----------|---------|----------|-------------|--------|--------|--------|--------|--|
| All Volumes                   |                              |         |           |         |          |             |        |        |        |        |  |
| Volume                        | Configuration                | Size    | Allocated | Free    | Capacity | Dedup Ratio | State  | Grow   | Export | Delete |  |
| zpool                         | raidz1 group: 1, devices: 12 | 8.69 TB | 1.41 MB   | 8.69 TB | 0%       | 1.00x       | ONLINE |        |        |        |  |
| Live Volumes Disks Statistics |                              |         |           |         |          |             |        |        |        |        |  |
| Disk                          | r/s                          | w/s     | kr/s      | kw/s    | wait     | actv        | wsvc_t | asvc_t | %w     | %b     |  |
| zpool (12 disks)              |                              |         |           |         |          |             |        |        |        |        |  |
| c0t50015178F35E20BDd0         | 2.8                          | 14.0    | 13.7      | 48.8    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E219Ad0         | 2.8                          | 14.0    | 13.7      | 48.8    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E2237d0         | 2.8                          | 14.4    | 13.7      | 50.4    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E2331d0         | 2.8                          | 14.2    | 13.7      | 49.6    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E236Dd0         | 2.8                          | 15.4    | 13.7      | 54.4    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E23D0d0         | 2.8                          | 15.0    | 13.7      | 52.0    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E2421d0         | 2.8                          | 17.4    | 13.7      | 62.4    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E247Ad0         | 2.8                          | 17.4    | 13.7      | 62.4    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E2513d0         | 2.8                          | 16.2    | 13.7      | 57.6    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E254Ad0         | 2.8                          | 16.2    | 13.7      | 57.6    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E26B2d0         | 2.8                          | 13.8    | 13.7      | 48.0    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |
| c0t50015178F35E910Bd0         | 2.8                          | 13.8    | 13.7      | 48.0    | 0.0      | 0.0         | 0.0    | 0.0    | 0      | 0      |  |

## 5.5 Configure NFS

Now that the storage has been configured it's time to share it, utilizing NFS. A total of 4 NFS mount points will be created, resulting in two mount points utilizing each 10GbE network port. In this design the 4 NFS mount points are named as NFS1a, NFS1b, NFS2a and NFS2b. This static load balancing will not be configured in the Nexenta\* appliance, but rather the systems utilizing the storage. When adding NFS storage to a VMware\* environment, the "a" mount points will utilize one 10GbE network port while the "b" mount points will use the other. This is accomplished by using the different IP addresses on the "a" and "b" network ports.

6. Create Folder
  - a. Navigate to the Folder field and select **Create**.

About | Support | Add Capacity | Register | Help

Status Settings Data Management Analytics

Data Sets Shares SCSI Target Auto Services Runners

Volumes

Show  
Summary Information

Create  
Create New Volume

Import  
Import Existing Volume

Folders

Show  
Summary Information

Create  
Create New Folder

Search  
Search

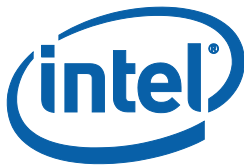
SUMMARY INFORMATION: VOLUMES

All Volumes

| Volume | Configuration                |
|--------|------------------------------|
| zpool  | raidz1 group: 1, devices: 12 |

Live Volumes Disks Statistics

| Disk                  | r/s | w/s |
|-----------------------|-----|-----|
| zpool (12 disks)      |     |     |
| c0t50015178F35E20BDd0 | 1.2 | 0.8 |
| c0t50015178F35E219Ad0 | 1.2 | 0.8 |
| c0t50015178F35E2237d0 | 1.2 | 0.8 |
| c0t50015178F35E2331d0 | 1.2 | 0.8 |
| c0t50015178F35E236Dd0 | 1.2 | 0.8 |
| c0t50015178F35E23D0d0 | 1.2 | 0.8 |



## Walk Through Setup and Configuration

- b. Select **zpool** Volume if it hasn't already been pre-populated.
- c. Specify the Folder Name and Description

Volume **zpool**  
Folder's volume.

Folder Name **NFS1a**  
Each folder pathname's component delimited by backslash ('/') can only contain alphanumeric characters and hyphens. Folder pathname must begin with an alphanumeric character and not end with a hyphen.

Description **NFS1a**  
Human-readable description for this folder.

- d. Select the default 128K for the Record Size.

Record Size **128K**  
Specifies a suggested block size for files in the folder. Default is 128K.

**Note:** The NFS record size denotes the largest record that can be used to accommodate files placed on the ZFS based storage. Smaller files will use record sizes closer to their size and larger files will do the same. For example, 3kb files will require a 4K record size, where as 1MB files would require 8x records at 128K. However, all these files will still reside in 4KiB ZFS sectors.

- e. Enable deduplication and select **OK** to the warning.
- f. Leave the remaining values at their defaults and select the **Create** button.

Deduplication **on**  
Controls the deduplication option for this dataset. If enabled, it will optimize storage by deduplicating identical data.

Compression **on**  
Controls the compression algorithm used for this dataset. Default is "lz4". Currently, "gzip" is equivalent to "gzip-6".

Number of Copies **1**  
Controls the number of copies of data stored for this dataset. Default is 1.

Case Sensitivity **mixed**  
Indicates whether the file name matching algorithm used by the file system and NFS at the same time. Default is "mixed".

The Folder Summary View is displayed, listing the newly created folder.

| SUMMARY INFORMATION : FOLDERS                   |           |           |         |                          |                          |                          |                          |                          |                          |                                     |
|-------------------------------------------------|-----------|-----------|---------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-------------------------------------|
| Folder                                          | Refer     | Used      | Avail   | CIFS                     | NFS                      | FTP                      | RSYNC                    | WebDAV                   | Index                    | Delete                              |
| <input checked="" type="checkbox"/> zpool/NFS1a | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| Results 1 - 1 (all)                             |           |           |         |                          |                          |                          |                          |                          |                          |                                     |

7. Create 3 additional Folders
  - a. Repeat Steps 1a – 1f specifying unique folder names.



## Walk Through Setup and Configuration

- b. When finished there will be a total of 4 folders listed.

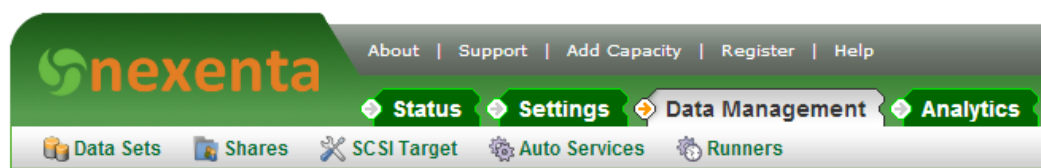
| SUMMARY INFORMATION : FOLDERS        |           |           |         |                          |                          |                          |                          |                          |                          |                          |                                     |
|--------------------------------------|-----------|-----------|---------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-------------------------------------|
| <input type="checkbox"/> Folder      | Refer     | Used      | Avail   | CIFS                     | NFS                      | FTP                      | RSYNC                    | WebDAV                   | Index                    | Delete                   |                                     |
| <input type="checkbox"/> zpool/NFS1a | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> zpool/NFS1b | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> zpool/NFS2a | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> zpool/NFS2b | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

Filter Delete selected Results 1 - 4 (all)

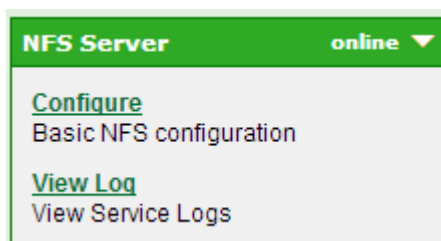
8. Enable NFS  
 a. Place a checkmark under NFS for each folder and select **OK** to the prompt.
9. Configure NFS

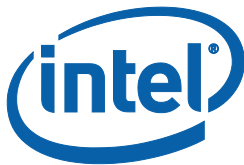
| SUMMARY INFORMATION : FOLDERS        |           |           |         |                          |                                          |                          |                          |                          |                          |                                     |                                     |
|--------------------------------------|-----------|-----------|---------|--------------------------|------------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-------------------------------------|-------------------------------------|
| <input type="checkbox"/> Folder      | Refer     | Used      | Avail   | CIFS                     | NFS                                      | FTP                      | RSYNC                    | WebDAV                   | Index                    | Delete                              |                                     |
| <input type="checkbox"/> zpool/NFS1a | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input checked="" type="checkbox"/> Edit | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> zpool/NFS1b | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input checked="" type="checkbox"/> Edit | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> zpool/NFS2a | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input checked="" type="checkbox"/> Edit | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> zpool/NFS2b | 256.00 KB | 256.00 KB | 7.60 TB | <input type="checkbox"/> | <input checked="" type="checkbox"/> Edit | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |

Filter Delete selected Results 1 - 4 (all)



- a. Navigate to Data Management and select **Shares**  
 b. Navigate to the NFS Server field and verify it is listed as online.  
 c. Select **Configure**





- d. Set the Client Version to 3<sup>◊</sup>.
- e. Set the Concurrent NFSD Servers to 2048 if not already.
- f. Leave the remaining values at their defaults and select the **Save** button.

**MANAGE NFS SERVER SETTINGS**

**Service State** ☐ Service is currently disabled  
Check it to enable service.

**Server Version** 4  
Sets the maximum version of the NFS protocol that will be registered and offered

**Client Version** 3  
Sets the maximum version of the NFS protocol that will be used by the NFS client attempted first. The default is 4.

**Concurrent NFSD Servers** 2048  
Maximum number of concurrent NFS requests.

**NFSD queue length** 64  
Connection queue length for the NFSD over a connection-oriented transport.

**Concurrent LOCKD Servers** 1024  
Maximum number of concurrent LOCKD requests.

**LOCKD queue length** 64  
Connection queue length for the LOCKD over a connection-oriented transport.

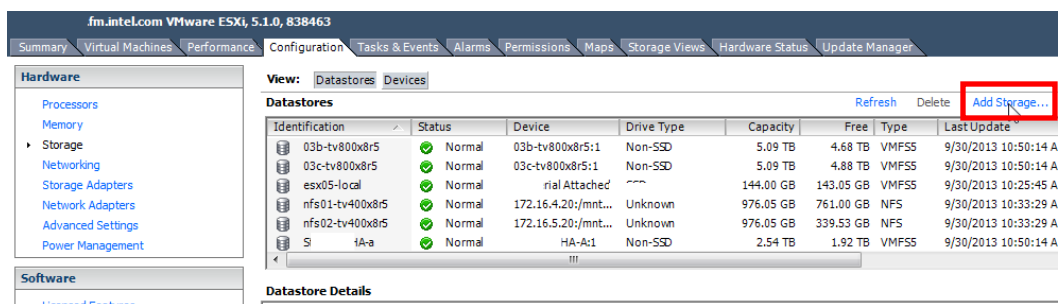
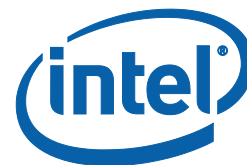
Save Restore defaults

<sup>◊</sup>At the time of this writing VMware vSphere\* does not support NFS version 4.

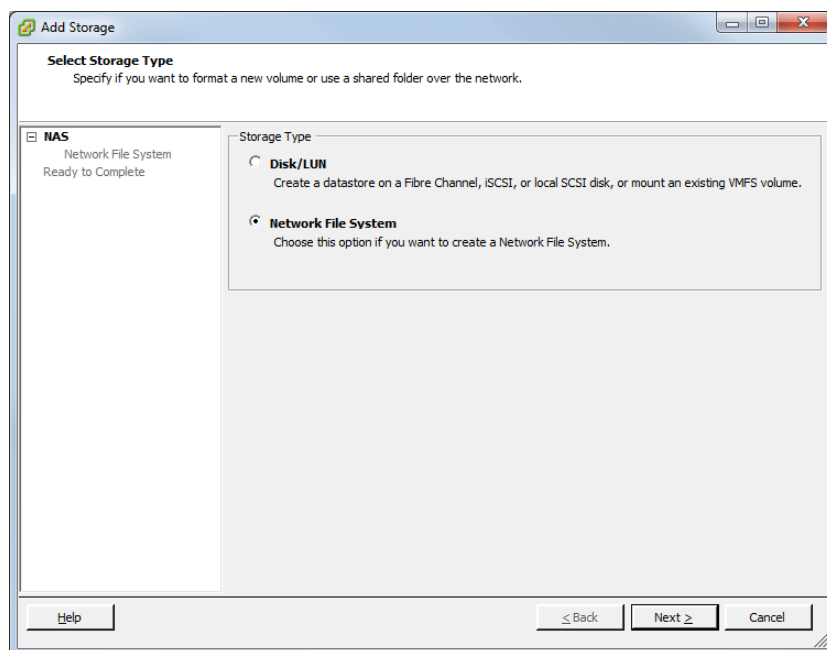
## 5.6 VMware\* NFS Configuration

After the NFS export configuration is complete on the Nexenta\* SW SAN:

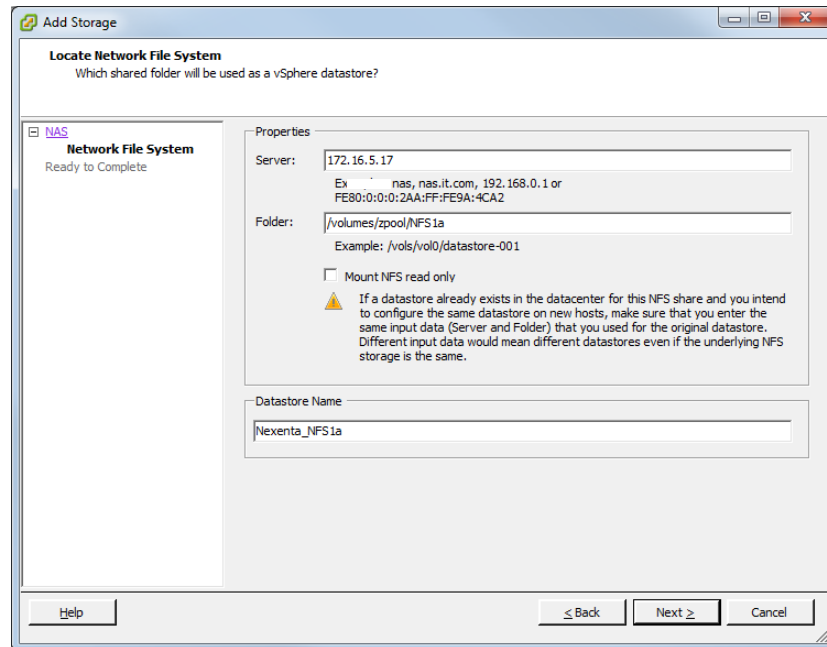
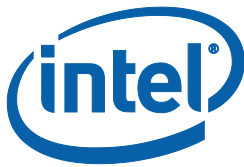
- 1) Open the VMware\*vSphere client console and add an NFS mount point in the VMware host. For our example we used the following private 172.16.x.x address and the mount point listed in the NexentaStor\* console we created in the previous steps.
- 2) On the **Configuration** tab of a selected host, click on **Add Storage...**



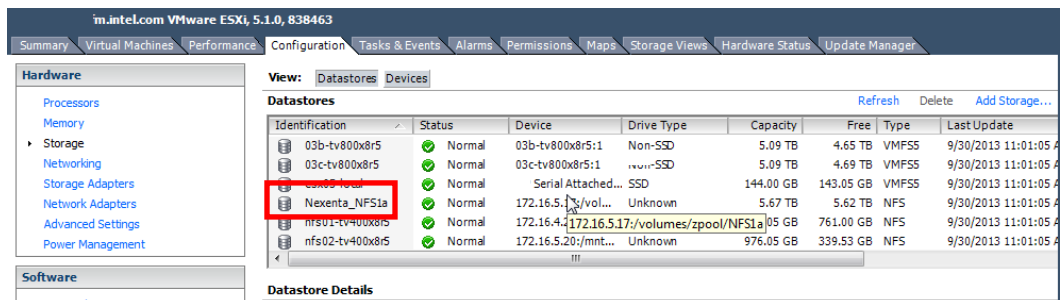
- Click the **Network File System** button if it is not selected, and click **Next**



- Follow the series of instructions for adding a network file system mount, IP address, and path.



Once this process is finished, you should end up with something similar to the screen shot below, which shows an NFS datastore in the VMware\* storage configuration tab. Repeat these steps for all the ESXi hosts which require access to the Nexenta\* SW SAN.



## 5.7 Calculating SSD Endurance in RAID

Although this endurance calculation method is similar to a ZFS Z1 set, ZFS uses stripes more efficiently and will likely result in a longer endurance than standard RAID5. When using RAID5 and SSDs, the endurance of the array = (N-1) where N is the number of disks in the RAID5 array, a RAID Z1 will have more endurance than a RAID5 set due to the variable stripe size & efficiency. Table 5-1 below shows an example calculation for a 400GB Intel® SSD in 5-disk RAID5 set. In Table 5-1, we use the formatted capacity of the drives in Bytes using base 2 (/1024) instead of base 10 (/1000). This calculation provides a more accurate representation of what an end-user will see in a formatted volume.

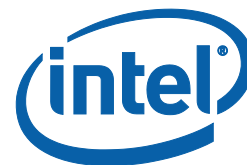


Table 5-1. Sample Endurance Calculation--5-Disk RAID5 Set DC S3500 Series

|                                                                                                             |               |
|-------------------------------------------------------------------------------------------------------------|---------------|
| Per Drive GB Usable Capacity                                                                                | 780GB         |
| Per Drive Overwrites/Day                                                                                    | .3 Overwrites |
| Per Drive Lifetime (years)                                                                                  | 5 Years       |
| Per Drive Lifetime (terabytes) =<br>Usable capacity X overwrites/day X days<br>X years/1024 (TB)/1024 (PB)  | 416 Terabytes |
| Number of disk in RAID set                                                                                  | 5 Disks       |
| <b>Equation:</b> Lifetime of 5 disk RAID5 = (5-1) *endurance (416TB)<br>or 1.6 Petabytes of write activity. |               |

## 5.8 Configuration and Setup Results

### 5.8.1 Results

- 4 hosts running against 4 NFS targets with 48 VMs @ 12VMs per NFS datastore
- Each VM configured with a 100% random 4K workload of 90% read/10% write
- Sustained 100K 100% Random IOPS across a 10GbE network at 2.5ms latency
- Power footprint of 650 watts, rack space of 2U, and estimated retail pricing under \$30K

### 5.8.2 Conclusion

The combination of Intel® Xeon® processors, Intel® x520-DA2 10GbE, and Intel® SSD DC S3500 Series with the Nexenta\* software based SAN using the methods described in this paper provides a high-IOPS solution capable of hosting hundreds of virtual machines. Given the endurance of the Intel® SSD DC S3500 Series, the solution presented here will last through 24-hour operation with a 10% write workload for 3-4 years using the sample endurance calculation outlined in the previous section. Finally, the small power and data center footprint plus low cost per IOPS when compared to traditional SAN supplies the business value required to pursue this non-traditional software based storage solution. With the price of SSDs continuing to decrease as storage capacity increases, the ongoing move toward the cloud and VMs, and increasing demand for cost-efficient IT services, the enterprise cloud community must look at alternative methods for supplying the data center with IOPS. Intel® SSDs and platforms will play an integral part in changing the face of storage in the data center.